a. In Fig. 13a I describe a "fixed (X, Y, Z) volume tracking camera" that may combine into Fig. 13b of a "non-scalable volume tracking matrix." I respectfully submit that this is exactly what Jain has taught for his camera arrangement.

b. In Fig. 14a I describe a "fixed (X, Y) area tracking camera" that may combine into Fig. 14b of a "scalable area tracking matrix." I respectfully submit that this is substantially and operatively different from Jain (i.e. Fig.'s 13a and 13b as well as Jain Fig. 2 and Fig. 5) for at least the following reasons:

- Jain's cameras are arranged "each at a different spatial perspective" to the scene and there is no teaching by Jain of either:
  - each camera's general orientation to the scene, e.g. substantially overhead and pointed down perpendicular to the object movement surface (i.e. the ground), or
  - each camera's relationship to its nearest neighbors, e.g. each camera's field-of-view is adjacent and overlapping to it's nearest neighbors.
- Both of these distinctions are critical and specified in my invention. They are the difference between Fig.'s 13a and 13b and Fig's 14a and 14b. Again, I respectfully submit that neither of these critical distinctions are taught or even indirectly employed by Jain.

c. Furthermore, with respect to Fig.'s 13a and 13b (i.e. Jain's type of camera arrangement), I specifically teach that:

- "Referring now to **Fig. 13a**, there is shown an example of a fixed (X, Y, Z) volume tracking camera ... It is important to note that, once in place, this volume tracking camera **126** has a fixed field-of-view (FOV) similar to a four-sided pyramid in shape within an image cone **121v**. The actual pixel resolution per inch of the FOV will vary throughout the height **121h** of the pyramid ranging from a higher value at the top width **121tw** to a lower value at the bottom width **121bw**. *These cameras are typically secured from an overhead position to have a perspective view arrangement **502m** of the desired tracking volume as shown in **Fig. 13b**.*" (italics added)
- "Referring now to **Fig. 13b**, there is shown one particular arrangement of fixed (X, Y, Z) cameras **502** that, when taken together, form a uniquely shaped tracking volume through which a player **17**... may transverse. *The resultant resolution per cross-sectional area of this volume **121tv** is non-uniform.* For example, while skating through any given point in the tracking volume... one body part of player **17** may be viewed by camera **126e** with a much lower resolution per inch than... a different body part. Also, the second camera such as **126d** may have a much different pixel resolution of marker **17sm** than camera **126e**. *Cameras **126a**, **126b** and **126c** may each have obstructed views of marker **17sm**.*" (italics added)
- *These are substantial problems.* Namely, the non-uniform, non-overlapping and perspective arrangement of multiple cameras creates a resulting distorted, disjointed and occlusion prone data set that is non-conducive to (X, Y) tracking of multiple small to large objects, such as in sports.

d. With respect to Fig.'s 14a and 14b (i.e. the arrangement I am claiming), I specifically teach that:

- "Referring now to **Fig. 14a**, there is shown an example of fixed (X, Y) area tracking camera... The entire assembly... *is preferably secured in an overhead position looking directly down at a subset of the tracking surface. From this overhead position, camera **124*** has a fixed FOV **120v** that is focused on the top surface of any players below and as such *maintains a substantially uniform pixel resolution per tracking area FOV **120v**.*" (italics added)
- "Referring now to **Fig. 14b**, *there is shown a scalable area tracking matrix **504m** comprising multiple fixed (X, Y) area tracking cameras **120c** aligned such that their FOVs **120v** are substantially side-by-side with a small overlap for calibration purposes.* Throughout this scalable matrix **504m**, the top surface **110** of player **17** can be readily tracked." (italics added)

- This combination of the an overhead position, looking directly down on the tracking surface as well as cameras which are substantially side-by-side with a small overlap are aspects of my invention that are specifically designed to overcome distortion, disjointed datasets and occlusions that are possible using solutions with camera arrangements such as Jain.

e. These points are summarized in the Operations section of my patent as following:

- "Due to its approach to camera placement, it is difficult to uniformly scale up to track larger areas such as a hockey rink." – i.e. Jain's approach;
- "Due to its strategy of fixed volume tracking accomplished with a complex overlapping network of camera field-of-views, the pixels resolution per player is inconsistent and the system is prone to lose markers when multiple players bunch up." – i.e. Jain's approach;
- "By limiting the fixed (X, Y) area tracking matrix to a top view only, the system creates a scalable approach to camera placement that provides a substantially uniform pixel resolution per area." – i.e. the teachings and claims of my patent;
- "By establishing a separate fixed (X, Y) area tracking matrix that continually locates each player in (X, Y) space the system provides the ability to automatically direct the pan, tilt and zoom aspects of one or more movable cameras to follow each player." – i.e. the teachings and claims of my patent;
  - (One should additionally note that Jain prefers no moving cameras.)

With respect to my second element of "a first algorithm… responsive to the fixed area tracking matrix for determining the (X, Y) location of each object" in comparison to Jain's second element of "a first algorithm responsive to each camera for determining the (X, Y) dimensional characteristics of each object and for forming a database per camera of each object's location, movements and dimensional characteristics" (again, my words) I believe that the following distinctions are also evident:

a. In my patent, the field-of-view orientation for each and every "area tracking camera" are co-planer and therefore the (X, Y) image data is more easily translatable into a single coplanar frame of reference that is essentially aligned with the playing field on which the objects are expected to travel. Hence, my patent teaches a highly scalable approach that allows for additional cameras to be placed in the matrix to maximally cover ever larger tracking areas, such that the potential object location simply expands laterally in both the X and Y dimensions without need of further translation;

b. For Jain, the fields-of-view for every camera are "each (pointed) at a different spatial location" (without necessary correlation), where the resulting "multiple two dimensional images of the real world scene" do not form a single coplanar frame of reference (see Col 8, Lines 49 – 66). While Jain could use his "world view" model to translate the (x, y) location of an object in a single field-of-view to the (X, Y) location of the object on the tracking surface, this is a more difficult task and as Jain puts in Col 8 Lines 66 though Col 9 Line 5:

- "The model is, it will be seen, not too hard to construct so long as there are, or are made to be, sufficient point of reference in the imaged scene. It is, conversely, almost impossible to construct the 3-D model, and select or synthesize the chosen image, of an amorphous scene";
  - Here, Jain is acknowledging that a camera freely oriented to the scene in general, and especially to the tracking surface, can result in a very complex field-of-view to calibrate – especially without pre-known markers in the field-of-view.

c. This difference is the one of the main reasons that I teach a first set of overhead cameras essentially coplanar to the tracking surface whose images are overlapping and pre-aligned so as to form a

single large composite image of consistent orientation and resolution through the resulting combined field-of-view. This approach obviates the need to require "sufficient point of reference in the imaged scene."

This difference also leads to the motivation to operate the "perspective view" cameras as a second "movable volume tracking" set, responsive to the (X, Y) object location information gathered by the first "fixed area tracking" set. With this distinct two-camera-set approach, the volume (i.e. 3D scene) tracking is made even more efficient because the perspective view cameras are now more easily calibrated and all available resolution is appropriately and maximally focused on the foreground (i.e. the moving players) rather than the fixed background.

With this in mind, and with respect to Jain's third element of "a second algorithm responsive to the databases formed by the first algorithm for combining the multiple two dimensional images and their accompanying object information, into a three-dimensional video database" (again, my words) I believe that the following distinctions are evident:

a. Jain is constructing his "three-dimensional video database" directly and exclusively from the original "non-alinged" two-dimensional images that are correlated via the "world-view" scene model.

b. My patent teaches that while a "top portion" of the tracked objects are located via the first, fixed "area tracking matrix," the remaining sides and bottom areas of the objects are picked up from the second, movable "volume tracking matrix."

c. And finally, Jain's stated purpose of the 3D video database is discussed in Col 9, Lines 52-61: "Second, in advanced embodiments of the system, the computer is not limited to selecting from the three-dimensional model a two-dimensional image this is, or that corresponds to, any of the images of the real-world scene as was imaged by any of the multiple video cameras. Instead, the computer may synthesize from the three-dimensional model a completely new two-dimensional image that is without exact equivalence to any of the images of the real-world scene as have been imaged by any of the multiple video cameras."

  - My first point would be to seriously challenge the very premise that each and every possible view (assuming it is not animated) could be created from these original perspective view cameras, especially since Jain has taken no care to at least specify a "360 degree" type coverage, with cameras placed every 6 degrees or so (which is the approach used by other companies when creating the "spinning" image effects such as in the super-bowl);

  - My second point is that Jain is inherently speaking about a different type of "3D video database" than I teach in my patent. Specifically:

    - With reference to at least Fig.'s 21a, 21b, 21c, 22a, 22b, 22c, 22d, 23a, 23b, 23c and 23d, my patent is teaching a better way of creating a 3D body-point model, which itself may then be used as a basis for rendering the game action from any desired view. As disclosed in the background to my invention, there are several systems currently collecting 3D body-point information which is ultimately a technique to support animation, not "natural video."

    - Jain is not collecting a body-point model in any classic sense, i.e. the list of centroids for the major body parts and joints. His system uses multiple 2D "natural video" images from various perspectives to *ideally* create a 360 degree "video-mesh" (my words) of the videoed activities. Using this 3D "video-mesh" filled in from only a few

perspectives, Jain postulates that you could then rotate your frame of reference and recreate the "natural look" from any virtual camera angle.

- As he points out numerous times in his application, the storage and processing requirements of the raw "video-mesh" *are significant* – especially in comparison to the minimal requirements of a classic body-point model.

- My third point would be that when Jain says that "Objects of interest in the scene are *identified…*" (Col. 8, Line 55 – italics added), he does not specifically teach how he intends to do this. In other words, how will his vision system automatically determine which play is "number 22 – Jones" vs. "number 17 - Hospodar?"
    - With respect to Fig.'s 21a, 21c and 24, I specifically teach the use of coded identity markers detectible by the image analysis software.

In summary, I respectfully submit that my teachings as claimed are substantially different from Jain, especially in light of each patent's own specification which further defines the common language used in both my and Jain's claims.

I thank you for your consideration in these matters.

Sincerely,

James A. Aman

This communication was faxed to (571)-273-7339 on 8/7/05. It was also mailed Post Office To Addressee from Harleysville, Pennsylvania on 8/08/05 using label 66968959.